DOCUMENT RESUME

ED 171 795 TM 009 440

AUTHOR Douglass, James B.
TITLE Generalizability of Behavior-Specific vs. General Items for Student Ratings of Instruction.
PUB DATE Apr 79
NOTE 18p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Analysis of Variance; *Behavior Rating Scales; Higher Education; Item Analysis; Mathematical Formulas; Rating Scales; Statistical Bias; *Student Evaluation of Teacher Performance; *Teacher Behavior; Test Items; *Test Reliability

ABSTRACT

Undergraduates in five classes were asked to rate their instructors on four general items, and on twenty-four behavior-specific items describing instructor involvement. Three reliability coefficients differing in error score assumptions were calculated for different numbers of students and items, to permit comparison of the reliabilities of both types of items. Each of the three coefficients was appropriate for a different situation. Internal consistency reliability was appropriate when no generalization was desired beyond a given set of students. Rater reliability was best when generalization beyond a given item or set of items was not needed. General reliability was appropriate when generalization was desired over both items and students. The general reliability of the behavior-specific items was higher than that of the general items, if there were relatively few students compared to items. Rater and internal consistency reliabilities exhibited the opposite results. The halo effect for the behavior-specific items was not significantly less than the general items, but the relationship was in the predicted direction. Finally, there was less leniency bias on the behavior-specific items, as predicted. (The twenty-eight items are appended). (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

James B. Douglass

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED171795

GENERALIZABILITY OF BEHAVIOR-SPECIFIC VS. GENERAL ITEMS FOR STUDENT RATINGS OF INSTRUCTION

JAMES B. DOUGLASS
Learning and Evaluation Service
Michigan State University

Student ratings can be used to improve instruction either directly or indirectly. Ratings can be used to give an instructor direct feedback from his/her students or they can be used to give administrators, peers or students information useful for promotion, salary or course selection decisions. Items that ask for general ratings of instruction may be appropriate for evaluations to be used by others but they are often of limited usefulness to an instructor seeking to improve his/her instruction. It would be more useful to have ratings of specific behaviors that relate to effective instruction. Items have been developed at the Learning and Evaluation Service to measure specific instructor behaviors. For each factor from Michigan State's general evaluation form nineteen to twenty-eight behavior-specific items have been developed. It is anticipated that instructors will investigate problem areas identified on the general form by administering the appropriate behavior-specific items. This will help instructors identify specific behaviors influencing low student evaluations.

The purpose of this study is to investigate properties of behavior-specific items in contrast to the more general evaluation items from which they were developed. A generalizability study will be conducted from which the effect of varying numbers of students and items can be determined for three different generalizability coefficients. The consistency of student response within classes, the halo effect and leniency bias across items and the relative magnitudes of item and student variances will also be investigated.

Related Research

In student rating of instruction the reliability of the mean rating of an instructor or class is of primary interest. Errors in mean rating may result from item sampling, student sampling or both item and student sampling. These multiple sources of error can be examined in the context of generalizability theory (Cronbach, Gleser, Nanda, and Rajaratnam, 1972). Differentially defining the components of true and error variance results in three useful reliability coefficients. If item sampling is considered a source of error and student sampling is not, a reliability coefficient that is a measure of the internal consistency of the items is obtained. The true score variance includes a component of class or instructor effect and a component of student variance. A coefficient of rater reliability is obtained when student sampling is considered a source of error and item sampling is not. The true score variance has a class effect component

Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, April 8-12, 1979.

Session 3.03

TM009 440

and an item component. If both items and students are considered to be sampled, the result is the most general coefficient with only a class effect component in the true score variance (Kane and Brennan, 1977).

Defining reliability as true score variance divided by the sum of true score variance and error score variance, it can easily be seen that for any given number of items and students the general coefficient will be less than or equal to both the item consistency and the rater reliability coefficients. The relative magnitudes of the item consistency and rater reliability coefficients depend upon the relative sizes of the item and student true score variance components and the relative numbers of items and students.

Given three reliability coefficients of different magnitudes the question of which is the most appropriate naturally arises. Ebel (1951) argues appropriately that between-raters or student variance should be included in the error score variance when different raters will rate different subjects. This is the situation in student rating of instruction at most colleges and universities where there is very little overlap of students between classes to be compared. Whether an item component should be included in the true score variance depends upon whether or not the given set of items is considered exhaustive of the population of items of interest. If the items are samples of possible items, it seems most reasonable to apply the most general reliability coefficient based on the assumption of both item and student sampling (Kane, Gillmore, and Crooks, 1976). Once the appropriate variance components have been calculated, each of the three coefficients can be estimated for any given number of items and students. Therefore the generalizability study can be performed without knowing which coefficient is most appropriate for any given situation.

The size of the general reliability coefficient increases as the item and student variance components decrease and as the number of items and students increase. It has been found that the item variance component is small compared to the student variance component so that reliability increases more rapidly by adding students than by adding items (Kane, Gillmore, and Crooks, 1976). Unfortunately it is much more difficult to add students to a class than it is to add items to a rating form. By reducing halo effect and increasing student agreement, behavior-specific items should increase the item variance and decrease the student variance component. As items are added the reliability should increase more rapidly for the behavior-specific items than for the general items.

The general evaluation items for this study are taken from the Michigan State University Instructional Rating System Form B (Appendix A-1). This form was developed from a previous form by changing the response options from strongly agree, agree, neither agree nor disagree, disagree and strongly disagree to superior, above average, average, below average, and inferior (Showers, 1974). The B form has less leniency bias than the previous form; however, the reduction in bias was not sufficient to significantly improve the rater reliability. Even if the rater reliability had increased it might have resulted from an increase in the size of the item component of the true score variance. This would have resulted in decreases in the item

consistency and general reliability coefficients. In this study the problem is recognized by investigating the effects of the various variance components upon all three reliability coefficients.

The behavior-specific items used in this study were derived from Form B which consists of twenty-one general evaluation items. The first twenty items are divided into five factors with four qeustions in each factor. In order, the factors are called Instructor Involvement, Student Interest, Student-Instructor Interaction, Course Demands and Course Organization. The twenty-first item is a general item with high loadings on all five factors. Olson (1978) prepared a separate form for each factor and had two to four classes of students respond to each form. Each form asked students to rate their instructor or class on each of the four general items and to state for each item what specific instructor actions or behaviors influenced their ratings. Appendix A-2 contains Olson's form for the Instructor Involvement factor. Responses to each question were collapsed into categories upon which behavior-specific items were based. Nineteen to twenty-eight behavior-specific items were developed to represent each factor. Balanced representation of a factor was considered more important than an exact correspondence between particular general evaluation items and particular behavior-specific items. For each factor a separate form was prepared containing the nineteen to twenty-eight behavior-specific items and the four general items. Appendix A-3 contains the items for the Instructor Involvement factor. Items developed for the other four factors can be found in Olson (1978).

Objectives

The four primary objectives are:

1. To estimate variance components useful for calculating reliability coefficients. Variance components will be estimated for general items and for the related behavior-specific items. Three generalizability coefficients differing in error score assumptions will be calculated for different numbers of students and items so that the reliabilities of the behavior-specific items can be compared to those of the general items. The three generalizability coefficients result from including an item variance component, a student variance component or both item and student variance components in the error score variance. The effect of the relative magnitudes of student and item variance upon the generalizability coefficients will be examined.

2. To test the hypothesis that variance across students is smaller for the behavior-specific items than for the general items. Since the behavior-specific items ask students to rate characteristics that are observable and specific, there should be more agreement among students on these items.

3. To test the hypothesis that variance across items is greater for behavior-specific items than for general items. Since the behavior-specific items require ratings of observable characteristics, there should be less halo effect.

4. To test the hypothesis that the mean of the general items is less than the mean of the behavior-specific items. The rating of observable characteristics should be less prone to leniency bias than the rating of general characteristics.

Method

Sample

Five undergraduate classes at Michigan State University were administered the Instructor Involvement behavior-specific and general items by their class instructors (Appendix A-3). Classes were chosen on the basis of convenience and included a business statistics class, two social science classes and two humanities classes. The smallest class had 32 complete sets of student responses. Student responses were randomly discarded in the other four classes until 32 complete sets remained in each.

Design

The design is a split-plot design labeled V-B by Cronbach (1972, p. 38), in which classes are crossed with items and equal numbers of students are nested within classes (see Figure 1). Each student responded to each of the behavior-specific and general items for the Instructor Involvement factor. The variance components were estimated separately for the behavior-specific and the general items allowing comparisons between the reliability coefficients of these two types of items.

		ITEMS	
		Behavior-Specific	General
		1 2 24	25 26 27 28
CLASSES	1	S_1	S_1
	2	S_2	S_2
	3	S_3	S_3
	4	S_4	S_4
	5	S_5	S_5

Figure 1: Split-plot design for analysis

Data and Instrumentation

The rating form contains instructions to the student, 24 behavior-specific items (1-24), and the four general items (25-28). The rating form is in Appendix A-3. The five instructors in the study administered the forms to their own classes. Students were asked to mark their responses on optical scan sheets. The data on the optical scan sheets was transferred to computer cards by the Michigan State University Scoring Office.

Model and Reliability Formulas

The model for the split-plot design is taken from Kane and Brennan (1977, pp. 271-2). The structural model is:

$$X_{csi} = \mu + \alpha_c + \pi_{s(c)} + \beta_i + \alpha\beta_{ci} + \beta\pi_{is(c)} + e_{o(csi)}, \quad (1)$$

where

μ = grand mean,

α_c = effect for class c ($c = 1, 2, \ldots, n_c$),

$\pi_{s(c)}$ = effect for student s ($s = 1, 2, \ldots, n_s$) nested within class c,

β_i = effect for item i ($i = 1, 2, \ldots, n_i$),

$\alpha\beta_{ci}$ = class by item interaction,

$\beta\pi_{is(c)}$ = item by person (nested within class) interaction, and

$e_{o(csi)}$ = experimental error (o is a replication subscript).

The analysis of variance table and formulas for calculation of the variance components based on this model are presented in Kane and Brennan (1977, p. 281).

The formulas for the three reliability coefficients are presented below. The general reliability coefficient with students and items sampled is:

$$\frac{\sigma^2(\alpha)}{[\sigma^2(\alpha)] + [\frac{\sigma^2(\pi)}{n_s} + \frac{\sigma^2(\alpha\beta)}{n_i} + \frac{\sigma^2_e}{n_i n_s}]} . \quad (2)$$

The rater reliability coefficient with students sampled and items fixed is:

$$\frac{\sigma^2(\alpha) + \frac{\sigma^2(\alpha\beta)}{n_i}}{[\sigma^2(\alpha) + \frac{\sigma^2(\alpha\beta)}{n_i}] + [\frac{\sigma^2(\pi)}{n_s} + \frac{\sigma^2_e}{n_i n_s}]} . \quad (3)$$

The internal consistency reliability coefficient with students fixed and items sampled is:

$$\frac{\sigma^2(\alpha) + \frac{\sigma^2(\pi)}{n_s}}{[\sigma^2(\alpha) + \frac{\alpha^2(\pi)}{n_s}] + [\frac{\sigma^2(\alpha\beta)}{n_i} + \frac{\sigma^2_e}{n_i n_s}]} . \quad (4)$$

The derivations of the reliability formulas are discussed in Kane and Brennan (1977, pp. 274-277).

Each coefficient can be interpreted as an expected true score variance divided by the expected observed score variance. Since the expected observed score variance is the same in each case, the coefficients differ only in expected true score variance. The differences depend upon whether the items or students in a particular study are considered to be populations or samples of populations. The item variance, $\sigma^2(\beta)$, does not appear in the expected score variance since the items are constant across all classes. If the items were not constant, $\sigma^2(\beta)$ would appear in the expected true score variance in formula (3) and in the expected observed score variance of all three formulas.

Procedure

The instructor involvement forms with 24 behavior-specific and four general evaluation items were given to the five class instructors. The instructors were also given scoring sheets, pencils and written directions. The directions indicated that each instructor was to ask his/her students to respond to every item on the form, and to return the forms and scoring sheets to the instructor without identifying themselves on the form or scoring sheet.

Thirty-two complete sets of responses were obtained in the smallest class. Student responses were randomly discarded in the other four classes until each contained 32 complete sets. This resulted in a balanced design with 32 students nested in each of five classes crossed with 28 items.

Separate analyses of variance were performed for the two sets of items. Estimated variance components were calculated for each set. The estimated variance components were used in formulas (2), (3), and (4) to calculate reliability coefficients.

Results

The estimated variance components for the behavior-specific and general items are presented in Table 1. Generalizability coefficients calculated from the variance components are presented in Table 2. The coefficients are given for samples of one and ten items and one and ten students. These numbers of items and students were chosen to give the reader an indication of how the coefficients vary as the number of items and students vary.

For the most general coefficient, the reliability tends to be higher for the behavior-specific items compared to the general items when there are relatively few students compared to the number of items. For 10 or more items and 10 or more students the values of the coefficient tend to be very similar for behavior-specific and general items.

The rater reliability is higher for the behavior-specific items since the estimated variance for the class by item interaction, $\hat{\sigma}^2(\alpha\beta)$, is larger for the behavior-specific items. The internal consistency reliability is higher for the general items since the estimated student variance, $\hat{\sigma}^2(\pi)$, is larger for the general items.

For either set of items the general reliability coefficient is always smaller than either the rater or internal consistency reliabilities since it contains only the estimated class effect, $\hat{\sigma}^2(\alpha)$, in the true score variance. The rater reliability will be smaller than the internal consistency reliability for the behavior-specific items whenever .11 divided by the number of items is less than .20 divided by the number of students. For the general items the rater reliability will be smaller than the internal consistency reliability whenever .07 divided by the number of items is less than .34 divided by the number of students. Thus for large numbers of students the rater reliability will tend to be higher than the internal consistency reliability for either set of items and for small numbers of students the internal consistency will tend to be higher. Notice that the internal consistency reliability is a particularly large overestimate of the general reliability for small numbers of students.

Table 1

Estimated Variance Components

Estimated Variance Component		Behavior-Specific Items	General Items
Class	$[\hat{\sigma}^2(\alpha)]$	.17	.21
Student	$[\hat{\sigma}^2(\pi)]$	.20	.34
Item	$[\hat{\sigma}^2(\beta)]$	.09	.06
Class X Item	$[\hat{\sigma}^2(\alpha\beta)]$	.11	.07
Error	$[\hat{\sigma}^2_e]$	.44	.33

Table 2

Estimated Values of Generalizability Coefficients

Generalizability Coefficients	$n_i = 1$		$n_i = 10$	
	$n_s = 1$	$n_s = 10$	$n_s = 1$	$n_s = 10$
General Reliability				
Behavior-Specific Items	.18	.49	.40	.83
General Items	.22	.59	.36	.83
Rater Reliability				
Behavior-Specific Items	.30	.81	.43	.88
General Items	.29	.78	.37	.85
Internal Consistency				
Behavior-Specific Items	.40	.55	.87	.93
General Items	.58	.68	.93	.96

Note: n_i = number of items; n_s = number of students.

The equalities of student variances and of item variances for the behavior-specific and general items were tested by comparing 95% confidence intervals. Each of the estimated variances is a difference of mean squares divided by a constant. Scheffé's procedure for estimating upper and lower bounds of confidence intervals for such variance components was used (1959, pp. 231-235).

The null hypothesis that student variances for behavior-specific and general items are equal is rejected at the .05 level of significance. The upper limit of the 95% confidence interval for the behavior-specific items is .25 and the lower limit of the 95% confidence interval for the general items is .26. The point estimates were .20 and .34 respectively. The research hypothesis that student variance is smaller for behavior-specific items is accepted.

The null hypothesis that item variances are equal for behavior-specific and general items is retained. The estimated variances are ordered in the predicted direction but failed to be significantly different. The point estimates are .09 and .06 respectively. The lower limit of the behavior-specific item variance is .05 and the upper limit of the general item variance is .11 with each at the 95% confidence level.

The hypothesis that the mean for the behavior-specific items is the same as the mean for the general items is rejected with $p < .001$. The t-test for equality of means for dependent samples results in $t = 6.67$ with $df = 159$. The means for each of the five classes and the overall means for the two types of items are presented in Table 3. The research hypothesis that the mean of the general items is less than the mean of the behavior-specific items is accepted.

Table 3

Class and Overall Means for
Behavior-Specific and General Items

Class	Behavior-Specific Items	General Items
1	2.80	2.60
2	1.70	1.30
3	2.40	2.31
4	1.99	1.84
5	2.04	1.88
Overall	2.18	1.99

$t_{overall} = 6.67$; $df = 159$; $p < .001$.

Discussion

Variance Components and Reliabilities

Each of the reliability coefficients previously discussed is appropriate for a different situation. The internal consistency reliability is appropriate when no generalization is desired beyond a given set of students. This is the least useful type of reliability for student ratings of instruction. Student variance is included in the true score variance although typically different students rate different classes and instructors. This is an example of not including rater variance in the error variance even though different raters are used for different ratings in actual practice (cf. Ebel, 1951). As well as being inappropriate conceptually, the internal consistency reliability is a large overestimate of the more appropriate general reliability. Although the internal consistency reliability is somewhat less for the behavior-specific items than the general items, it is still an unacceptable estimate of the general reliability.

The rater reliability is appropriate when there is no desire to generalize beyond a given item or set of items. Although this is relatively rare in student rating of instruction, one situation appropriately involving rater reliability does arise for behavior-specific items. This situation occurs when an instructor is interested in students' ratings of a particular behavior rather than a general characteristic. In this situation it is appropriate to look at the rater reliability of a single behavior-specific item. For the behavior-specific items in this study the rater reliability is quite high. For any one item the rater reliability for 10, 30 and 60 students is .81, .93, and .96, respectively.

The most general reliability is appropriate when generalization is desired over both items and students. This is typically the situation in student rating of instruction. Different students usually rate different instructors. And although the same items may be used to rate different instructors, the items are conceptually samples of possible items.

The behavior-specific items have higher general reliabilities than the general items when there are few students compared to the number of items. Thus behavior-specific items would be more reliable than general items for small classes. For small classes the reliability of a rating form can be increased more rapidly by adding behavior-specific items than by adding general items. This occurs because the ratio of the item variance component, $\hat{\sigma}^2(\alpha\beta)$, to the student variance component, $\hat{\sigma}^2(\pi)$, is .55 for the behavior-specific items compared to .21 for the general items [see (2)]. Behavior-specific items thus help solve the dilemma, pointed out by Kane, Gillmore and Crooks (1976), that it is more efficient to increase reliability by adding students rather than items. Since the problem only occurs for small numbers of students, the addition of behavior-specific items will more efficiently increase the reliability. Since the general reliabilities are very similar when there are 10 or more items and 10 or more students, for larger classes reliability should not be a consideration in determining item type for forms with 10 or more items.

Student Agreement

It was anticipated that the student variance for the behavior-specific items would be smaller than the student variance for the general items. Students should be able to agree with one another more when they are evaluating observable behaviors than when they are evaluating inferred attitudes or characteristics. Although this hypothesis was supported by the data, two caveats should be noted. First the student variance is confounded with the student by class interaction because of the nested design used in this study. There is the possibility that the difference is caused at least in part by the student by class interaction. It is worth noting that this does not affect the values of the reliability coefficients since the confounded terms would always appear together in generalizability coefficients estimated from a fully crossed design.

The second caveat is that although the statistical test assumes independent samples, the samples are dependent. Unfortunately no appropriate test exists for dependent samples. The test used is conservative in the sense that comparing 95% confidence intervals for the two parameters is a two-tailed test. Since the hypothesis was directional, a one-tailed test would have been justified. The dependent samples also lead to a more conservative test. The conservatism of the statistical test helps compensate for the influence of the confounded terms on the alpha level.

In subsequent studies the variance components for the general and behavior-specific items should be estimated from different samples of students to avoid the problem of dependent samples. In addition, students should be crossed with classes to avoid confounding the student variance with the student by class interaction.

Halo Effect

It was expected that the more specific and concrete behavior-specific items would have more item variance than the general items. Students should rate behavior-specific items without as much of a tendency to rely on general impressions of the instructor or class. Although the item variance is half again as large for the behavior-specific items, the difference fails to be statistically significant. There are at least four possible explanations for a lack of statistical significance even if a difference exists.

1) As discussed above, the statistical test is conservative.

2) There are only four general items on the form. The confidence interval for the general items is very large because an F value with 3 as one of its degrees of freedom greatly affects the size of the interval. This problem could be avoided in future studies by increasing the number of general items.

3) The estimated item variances are very small. This probably occurs because only one factor is represented on the form. It would be more

appropriate to compare the item variances of behavior-specific items with general items when each of the five factors is represented by both behavior-specific and general items.

4) Some of the behavior-specific items may be fairly general. Behavior-specific and general items form a continuum rather than a dichotomy and differentiation is a matter of judgement. In future studies groups of judges could be used to rate items on degree of specificity. The ratings could be used to dichotomize the items.

Leniency Effect

There has been a consistent leniency effect for the general items (Showers, 1974). For example, although the midpoint of the numerical scale is 3, the mean for the four general items was 2.12 in the Fall term of 1977. This mean is based on a sample of 387 classes with a total of 15,662 students. It was expected that the more observable characteristics measured by the behavior-specific items would result in more objective ratings and hence a mean closer to 3.0. This result did occur despite the fact that the leniency bias of the general items had been previously reduced by the introduction of the superior to inferior rating scale (Showers, 1974).

Conclusions

The following statements summarize the results and implications of this study for the use of behavior-specific and general items for student rating of instruction.

1. The general reliability of the behavior-specific items is higher than the general reliability of the general items if there are relatively few students compared to items.

2. The general reliability is .83 when there are 10 items and 10 students for either set of items.

3. There is no practical difference in general reliabilities with 10 or more items and 10 or more students.

4. The rater reliability of the behavior-specific items tends to be higher than the rater reliability of the general items.

5. The internal consistency reliability of the general items tends to be higher than the internal consistency of the behavior-specific items.

6. The rater reliability is a large overestimate of the general reliability when there are few items of either type.

7. The internal consistency reliability is a large overestimate of the general reliability when there are few students.

8. For a single behavior-specific item the rater reliability is .81 for 10 students.

9. There is more student agreement on ratings for the behavior-specific items.

10. There is not significantly less halo effect for the behavior-specific items, although the relationship is in the predicted direction.

11. There is less leniency effect for the behavior-specific items.

References

Cronbach, L. J., Gleser, G. C., Nanda, H., and Rajaratnam, N. The dependability of behavioral measurements: Theory of generalizability for scores and profiles. New York: Wiley, 1972.

Ebel, R. L. Estimation of the reliability of ratings. Psychometrika, 1951, 16, 407-424.

Kane, M. T. and Brennan, R. L. The generalizability of class means. Review of Educational Research, 1977, 47, 267-292.

Kane, M. T., Gillmore, G. M., and Crooks, T. J. Student evaluations of teaching: The generalizability of class means. Journal of Educational Measurement, 1976, 13, 171-183.

Olson, L. A. Behavior-specific items for student rating of instruction. Paper presented at the meeting of the American Educational Research Association, Toronto, March 1978.

Scheffé, H. The analysis of variance. New York: Wiley, 1959.

Showers, B. Student instructional rating system: The effects of three kinds of response options on student ratings of instruction. Office of Evaluation Services, Michigan State University, 1974.

A-1

B

MICHIGAN STATE UNIVERSITY

STUDENT INSTRUCTIONAL RATING SYSTEM FORM
Form B

One way in which a teacher can improve his or her class is through thoughtful student reactions. This instructor hopes to use your responses for self-examination and self-improvement.

Please omit any of the items which do not pertain to the course that you are rating. For example, if you have had no homework assignments in this course omit (leave blank) those items pertaining to homework. With a pencil respond to the items using the KEY.

S - SUPERIOR: exceptionally good course or instructor
AA - ABOVE AVERAGE: better than the typical course or instructor
AV - AVERAGE: typical of courses or instructors
BA - BELOW AVERAGE: not as good as the typical course or instructor
I - INFERIOR: exceptionally poor course or instructor

	KEY	S	AA	AV	BA	I
1. The instructor's enthusiasm when presenting course material	1.	S	AA	AV	BA	I
2. The instructor's interest in teaching	2.	S	AA	AV	BA	I
3. The instructor's use of examples or personal experiences to help get points across in class	3.	S	AA	AV	BA	I
4. The instructor's concern with whether the students learned the material	4.	S	AA	AV	BA	I
5. Your interest in learning the course material	5.	S	AA	AV	BA	I
6. Your general attentiveness in class	6.	S	AA	AV	BA	I
7. The course as an intellectual challenge	7.	S	AA	AV	BA	I
8. Improvement in your competence in this area due to this course	8.	S	AA	AV	BA	I
9. The instructor's encouragement to students to express opinions	9.	S	AA	AV	BA	I
10. The instructor's receptiveness to new ideas and others' viewpoints	10.	S	AA	AV	BA	I
11. The student's opportunity to ask questions	11.	S	AA	AV	BA	I
12. The instructor's stimulation of class discussion	12.	S	AA	AV	BA	I
13. The appropriateness of the amount of material the instructor attempted to cover	13.	S	AA	AV	BA	I
14. The appropriateness of the pace at which the instructor attempted to cover the material	14.	S	AA	AV	BA	I
15. The contribution of homework assignments to your understanding of the course material relative to the amount of time required	15.	S	AA	AV	BA	I
16. The appropriateness of the difficulty of assigned reading topics	16.	S	AA	AV	BA	I
17. The instructor's ability to relate the course concepts in a systematic manner	17.	S	AA	AV	BA	I
18. The course organization	18.	S	AA	AV	BA	I
19. The ease of taking notes on the instructor's presentation	19.	S	AA	AV	BA	I
20. The adequacy of the outlined direction of the course	20.	S	AA	AV	BA	I
21. Your general enjoyment of the course	21.	S	AA	AV	BA	I
STUDENT BACKGROUND: Select the most appropriate alternative.						
22. Was this course required in your degree program?	22.	yes	no			
23. What is your sex?	23.	M	F			
24. What is your overall GPA? (a) 1.9 or less (b) 2.0-2.2 (c) 2.3-2.7 (d) 2.8-3.3 (e) 3.4-4.0	24.	a	b	c	d	e
25. What is your class level? (a) Freshman (b) Sophomore (c) Junior (d) Senior (e) Graduate or other	25.	a	b	c	d	e

Optional Items: Please answer any optional items according to the directions provided.

26	A	B	C	D	E	27	A	B	C	D	E
28	A	B	C	D	E	29	A	B	C	D	E
30	A	B	C	D	E	31	A	B	C	D	E
32	A	B	C	D	E	33	A	B	C	D	E
34	A	B	C	D	E	35	A	B	C	D	E
36	A	B	C	D	E	37	A	B	C	D	E
38	A	B	C	D	E	39	A	B	C	D	E
40	A	B	C	D	E	41	A	B	C	D	E
42	A	B	C	D	E	43	A	B	C	D	E
44	A	B	C	D	E	45	A	B	C	D	E
46	A	B	C	D	E	47	A	B	C	D	E
48	A	B	C	D	E	49	A	B	C	D	E
50	A	B	C	D	E	51	A	B	C	D	E

O-9640

STUDENT INSTRUCTIONAL RATING SYSTEM FORM (Written Comments)

One way in which an instructor can improve his class is through thoughtful student reactions. This instructor hopes to use your responses for self-examination and self-improvement. If you have any comments to make concerning the instructor or the course, please write them in the shaded area below.

MICHIGAN STATE UNIVERSITY

FORM B 1

STUDENT INSTRUCTIONAL RATING SYSTEM

One way in which a teacher can improve his or her class is through student reactions. This instructor hopes to use your responses for self-examination and self-improvement.

Respond to each of the four items below by circling the appropriate response to the right of the item. Then list illustrative specific instructor behaviors as requested.

KEY: S - SUPERIOR: exceptionally good instructor
AA - ABOVE AVERAGE: better than the typical instructor
AV - AVERAGE: typical instructor
BA - BELOW AVERAGE: worse than the typical instructor
I - INFERIOR: exceptionally poor instructor

EXAMPLE: The instructor's encouragement to students to express opinions S AA AV (BA) I
What specific instructor behaviors or actions influenced your rating on this item?
The instructor ignored students' opinions

1. The instructor's enthusiasm when presenting course material S AA AV BA I
What specific instructor behaviors or actions influenced your rating on this item?

2. The instructor's interest in teaching S AA AV B I
What specific instructor behaviors or actions influenced your rating on this item?

3. The instructor's use of examples or personal experiences to help get points across in class S AA AV BA I
What specific instructor behaviors or actions influenced your rating on this item?

4. The instructor's concern with whether students learned the material S AA AV BA I
What specific instructor behaviors or actions influenced your rating on this item?

Michigan State University

Student Instructional Rating System

FORM B1

One way in which an instructor can improve his or her class is through thoughtful student reaction. This instructor hopes to use your responses for self-examination and self-improvement.

Using the pencil provided, mark the appropriate space on the answer sheet according to the following scale:

Please do not put your name or student number on the answer sheet.

1. Superior: exceptionally good instructor
2. Above Average: better than the typical instructor
3. Average: typical instructor
4. Below Average: worse than the typical instructor
5. Inferior: exceptionally poor instructor

The instructor's:

1. Repetition of important points
2. Use of vivid examples
3. Intensity in presenting material
4. Use of voice modulation (changes)
5. Encouragement of students to go beyond class material
6. Use of facial expressions
7. Current knowledge of the subject
8. Excitement when teaching
9. Confidence in front of class
10. Use of examples from personal experience
11. Use of new material
12. Use of full class period for instruction
13. Use of examples to which you can relate
14. Use of audio-visual aids to illustrate points
15. Use of examples to emphasize important points
16. Use of an appropriate number of examples
17. Use of the chalkboard
18. Coverage of required reading in class
19. Encouragement of students to come in during office hours
20. Clarity of explanations
21. Pace of class presentation
22. Test questions or problems
23. Promptness in beginning class
24. Use of humor
25. Enthusiasm when presenting course material
26. Interest in teaching
27. Use of examples or personal experiences to help get points across in class
28. Concern with whether students learned the material